iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30796

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Convolutional Neural Network-VGG16 for Road Extraction from Remotely Sensed Images

Prajakta Ganakwar[1], Ms. Saroj Date[2]
[1, 2]*Department of Computer Science & Engineering, MGM's Jawaharlal Nehru Engineering College, Aurangabad*

Abstract: ***This study is performed to analyze the use of VGG16 in providing and improving the road extraction from remote sensing images (RSIs). As you can see in the past few years deep CNN is widely used in various applications to detect patterns and to analyze them as well. VGG16 which is also known as oxfordNet is a convolutional neural network (CNN). VGG16 mainly serves the purpose of classification and detection of objects. U-Net is considered as one of the standard CNN architectures for image classification tasks. It is considered as a best network for fast and precise segmentation of images. In this paper, we address the issue of speed and size by proposing a compressed convolutional neural network model namely Residual Squeeze VGG16. Proposed model compresses the earlier very successful VGG16 network and further improves on following aspects: (1) small model size, (2) faster speed, (3) uses residual learning for faster convergence, better generalization, and solves the issue of degradation, (4) matches the recognition accuracy***.

Keywords*: Road extraction, Remote sensing images, Convolutional neural Network (CNN), VGG16.*

I. INTRODUCTION

In the domain of pattern recognition and remote sensing images (RSI) detecting a pattern or object has been vibrating topic of all the times. Airports, Roads, Forests, Buildings, and urban settlements are often been the area of interest in RSI, with Roads being highly used for the purpose of navigation on lands and maps. In recent years it has been applied in many domains, e.g., urban planning, Geographic Information (GIS) data updating, and traffic navigation. Since roads have obvious geographical features, they have regular shapes in object extraction [1,4], with stripe-like distribution, geometric shapes of fixed width, and interconnected network topologies. So, the main problems been faced while road detection are as follows: Diversity. There are various types of roads present in the world. For example, highways, urban trunk roads, and country roads, resulting in multiscale characteristics. Narrowness. In comparison with huge tall buildings roads appear narrow, likely to cause discontinuous extraction. Easily disturbed. Trees present on roads can easily obscured the texture of the road or can confused with rivers in remote sensing images which leads to feature variation in different imaging conditions. Therefore, extracting roads from remote sensing images automatically and precisely is rather tough work. The road network always have a standard geometrical morphology, but extraction of road network from satellite images is not so easy as road network are usually covered by ground objects likes trees, vehicles and shadows. Therefore the color and shape of the road are different at different areas [2]. In this proposed method, VGG16 is used for extracting roads from remote sensing images. And CNN as classifier and gives more accuracy hence it is used. Convolutional neural network (CNN), which is an extension of biologically inspired multi-layer perceptron's (MLPs) [10], is considered as an one of the efficient image feature extractor. CNN architecture includes number of convolution as well as pooling layers, which are followed by again fully connected layers. The convolution layers includes local filters, that are adjust during the training of the CNN architecture to exploit the strong spatial local correlation present in the input images. CNNs are considered to be a leading technique in image classification and that have state-of-the-art performance in various applications includes handwritten digit recognition, traffic signs classification [12], and 1000 classImage Net dataset classification and localization.[10] So, this study gives idea about a deep extraction technique which inspired from the existing efficient techniques. CNN-based road extraction system, adopting a two-stage road.

II. LITERATURE SURVEY

A convolutional neural network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. Layers in CNN operate on local input regions, which we called them receptive fields. Receptive fields, parameter sharing and spatial subsampling are characteristics of CNN. The lack ofdependence on prior knowledge is a major advantage for CNN. Because the high degree of invariance of image distortion, CNN features usually have a better performance than the human-designed features. Convolutional neural network is a type of feed-forward artificial neural network, which can extract features for visual recognition tasks automatically.

Experiments have proved that the features extracted by CNN directly have better performances than the features extracted by human experience. Thus, we aim to use CNN to extract features and detect road candidates.[9,13]

Deep learning has given way to a new era of machine learning, apart from computer vision. Convolutional neural networks have been implemented in image classification, segmentation and object detection. Despite recent advancements, we are still in the very early stages and have yet to settle on best practices for network architecture in terms of deep design, small in size and a short training time. In this paper, we address the issue of speed and size by proposing a compressed convolutional neural network model namely CNN VGG16.

III.METHODOLOGY

Every proposed method is made up of collection of modules, this method has mainely three modules that are Acquiring dataset, tanning the dataset using VGG16, and last is testing that has been done using web application created using python. Different sets of images has been used for training and testing purpose.

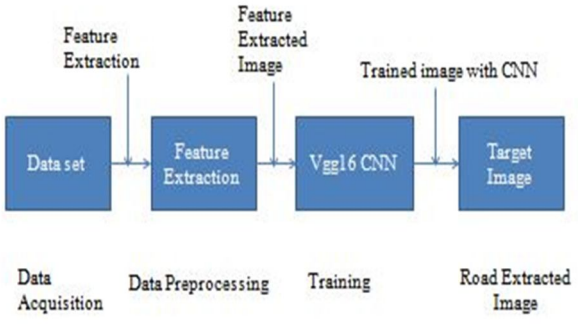


Figure 1: Proposed Architecture

A. Convolutional Neural Network

CNN has always been the most prefered way for fature extraction, for classification of images, accurate object detection as well as for segmentation too. Thus, here we have tried to work on cnn along with VGG16. To extracting the road features from high resolution satelite images. CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.

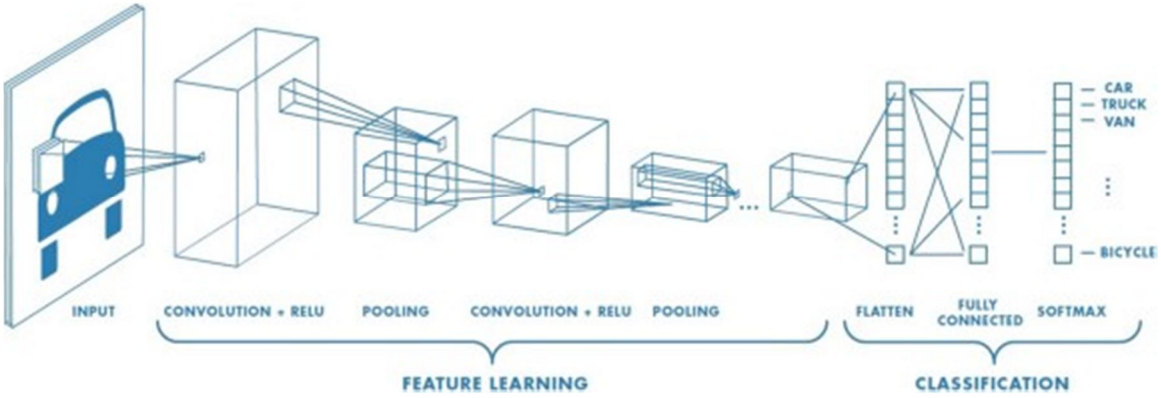


Figure 2: Neural network with many convolutional layers.

B. Convolution Layer

Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters.

C. Pooling Layer

Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or downsampling which reduces the dimensionality of each map but retains important information. Spatial pooling can be of different types:

1) Max Pooling
2) Average Pooling
3) Sum Pooling

Max pooling takes the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

D. Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like a neural network. With the fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as softmax or sigmoid to classify the outputs as cat, dog, car, truck etc.

E. VGG16

Deep learning has given way to a new era of machine learning, apart from computer vision. Convolutional neural networks have been implemented in image classification, segmentation and object detection. Despite recent advancements, we are still in the very early stages and have yet to settle on best practices for networkarchitecture in terms of deep design, small in size and a short training time. Convolutional Neural Network (CNN) is one of the latest classification methods proposed in this study using the famous VGG16 architecture. VGG16 training can last a long time if trained with random initialization of weights. For this reason, we selected initial weights by transfer learning to improve accuracy and speed up training time.[9]

IV. EXPERIMENTAL SETUP

This section discusses about following things:

A. Architecture of the CNN

The input given to CNN was a fixed size is of 227×227×3 pixel color image. The mean color image, determine on training color images, which is subtracted from each pixel. The Convolutional Neural Network model has five convolution layers and three fully connected layers. The first convolution layer hire 64 filters consists of size 11×11. The convolution stride is having total four pixels. The rectification linear unit (RELU) and local response normalization layers follow the first and second convolution layers. The architecture having total five max-pooling layers, which follow some of the convolution layers. A 3×3 pixel window, with stride 2 is used to perform pooling operation. The second convolution layer filters the output of the previous layer by using 256 filters of size 5×5. The convolution stride is one pixel and spatial padding is two pixels. The third convolution layer also hire 256 filters of size 3×3. The convolution stride and spatial padding are one pixel. A ReLU layer follows the third convolution layer. The fourth and fifth convolution layers have the same structure as the third convolution layer. As mentioned earlier, three fully connected layers follow the convolution layers. Each fully connected layer has 4096 nodes. Dropout probability is selected as 0.5. A loss layer is used as the last layer.[2]

B. Architecture of the VGG16

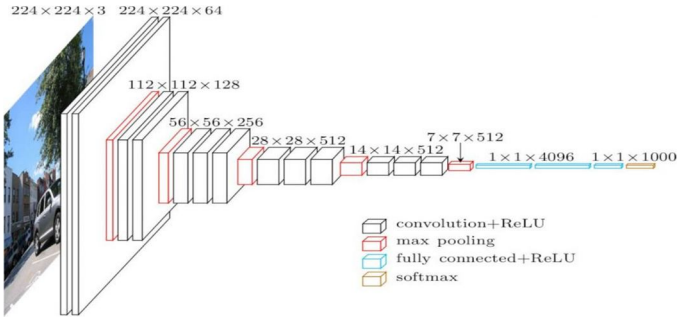


Figure 3: VGG16 Architecture

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU's.

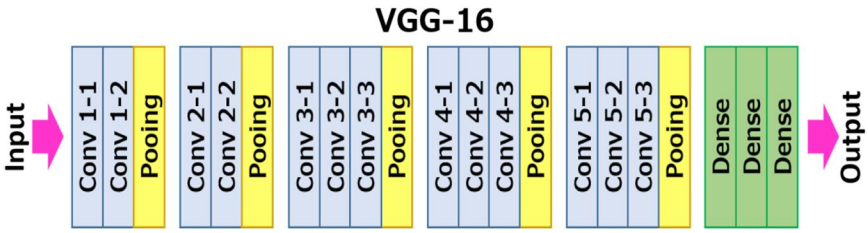


Figure 4: VGG16 Structure.

The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.[12]

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks. All hidden layers are equipped with the rectification (ReLU) non-linearity. It is also noted that none of the networks (except for one) contain Local Response Normalisation (LRN), such normalization does not improve the performance on the ILSVRC dataset, but leads to increased memory consumption and computation time.

C. Configurations

The ConvNet configurations are outlined in figure 4. The nets are referred to their names (A-E). All configurations follow the generic design present in architecture and differ only in the depth: from 11 weight layers in the network A (8 conv. and 3 FC layers) to 19 weight layers in the network E (16 conv. and 3 FC layers). The width of conv. layers (the number of channels) is rather small, starting from 64 in the first layer and then increasing by a factor of 2 after each max-pooling layer, until it reaches 512.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 5: Configuration of VGG16.

D. Dataset Construction

Many experiments were applied to calculate the performance of the proposed system on real remote sensing images which obtained from Massachusetts Roads dataset. A total of 1500 images were in collection, and from that all were used for training set and 44 different images have been used in testing set. The various preprocessing operations are performed to preprocess the image. Then applying Convolutional Neural Network with VGG16 on extracting the roads from the provided images.

E. Dataset Augmentation

The training dataset was artificially enlarged to reduce over fitting. Data set has been enlarged using augmentation strategy so that results can be improved. Various rotation operations has been made on input images along with horizontal as well as vertical flipping, which helped to enlarge the quantity of dataset that eventually leaded to improving the results.

F. Evaluation Matrix

To quantitatively evaluate the performance of different frameworks on road extraction we use two common metrics, as shown in equations below which are precision, recall. All of them are based on four basic components in information retrieval: true positive (TP), true negative (TN), false positive (FP), and False negative (FN). TP indicates the number of correctly classified road pixels, TN indicates the number of correctly classified background pixels, FP indicates the number of mistakenly classified background pixels, and FN indicates the number of mistakenly classified background pixels.

$$Precision = \frac{TP}{TP+FP}$$
$$Recall = \frac{TP}{TP+FN}$$

Precision is the ratio of correctly classified road pixels among all predicted road pixels, while recall measures the percentage of correctly classified road pixels among all actual road pixels. The higher the value, the better the performance.

V. RESULT

To practically implement the proposed method, Pycharm 2018 has been used. Total 1500 images were taken for training and 44 were used for testing purpose. After training the images using CNN and VGG16, which includes multi-layer network operations, it gives the output as extracted roads. Following figures, figure no. 5 to 8, shows the input and output obtained after successful implementation of the project. Here we are giving some sample images.

In every sample, input is given as one remotely sensed Image and as an output it identifies the roads from that input and displays it. Figure 6 has a roads in highly dense region, where as Figure 7 has a large part of road. Whereas Figure 8 and Figure 9 shows the successfully extracted roads, from the given input images.

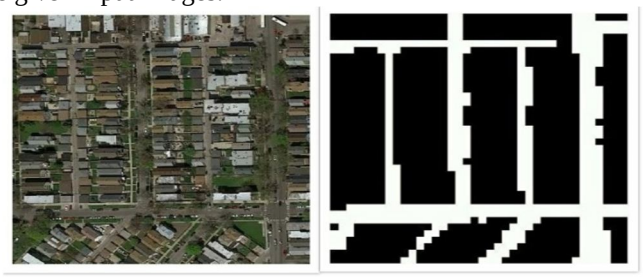


Figure 6: Sample 1: Input & Output

Figure 7: Sample 2: Input & Output 2

Figure 8: Sample 3: Input & Output

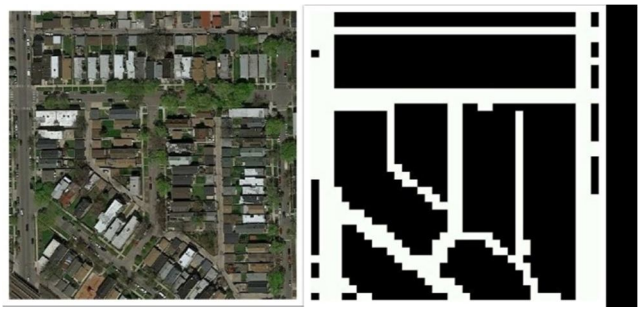


Figure 9: Sample 4: Input & Output 4

To evaluate the overall performance of the proposed scheme on the test set, the performance measures used are as accuracy, Precision and Recall. The results are given in Table 1.

Table 1. Performance Comparison

Method	Accuracy	Precision	Recall
K-means	92.72%	75.15%	74.2%
CNN	90.9%		
MFPN	82.1%	85.1%	71.2%
CNN VGG16	97.08%	80.66%	73.00%

As shown in Table 1, the CNN VGG16 model obtained 97.08% Accuracy, 80.66% Precision and 73% Recall, that parameters are measured using the no of roads classified correctly or incorrectly. These results show that the CNN VGG16 model successfully classified both roads and other regions. The performance of the proposed method was also compared to that of another method described in Refs. [5] that are image processing using edge detection and K- means .This method was chosen because of their successful result and methodological similarity. So as compared to image processing, proposed method was found more effective in feature extraction and classification stage. The convolution kernels were more effective in extracting the features that discriminated road and other regions. However, the image k-means features were encoded for the verification stage.

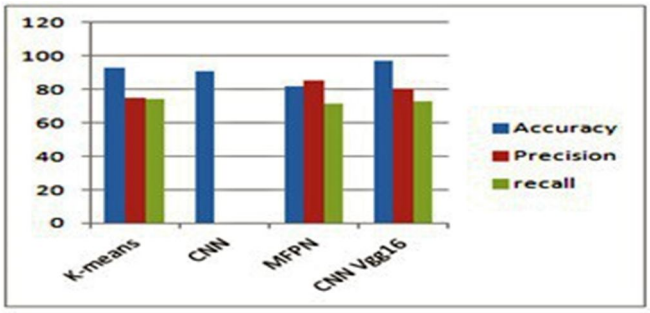


Figure 10: Graphical comparison

Above Figure 10 shows ,the Graphical representation of the result comparison.

VI. CONCLUSION

In this paper, the use of a deep CNN with VGG16 for the road extraction was analyzed. After potential roads extracted, the CNN architecture was trained for detecting roads on satellite images. So the proposed method successfully detected the various road parameters on the remote sensing images and gives more accuracy than previously used methods as K-means.

In this paper, we have built a CNN VGG16 model for road extraction from high resolution remote sensing images. Firstly we have extracted the features and then classified the images using CNN VGG16 classifier. Training on the data set has been for about 24 hours. After training the code we got our desired output image. The experiment was performed on Massachusetts data set and it showed better results as compare to other techniques.

However, we still think that the quality of an output image can be improved further along with less training time.

REFERENCES

[1] Xun Gao, Xian Sun, Yi Zhang, Menglong Yan, & Guangluan Xu,. "An End-to-end neural Network for Road Extraction from Remote sensing imagery by Multiple feature Pyramid Netwoek", IEEE Access, 2019.

[2] Yue Li, Rong Zhang and Yunfei Wu1," Road Network Extraction In High-Resolution Sar Images Based Cnn", 2017 Ieee International Geoscience And Remote Sensing Symposium (Igarss), pp. 1-4.

[3] Priya Singh, Ratnakar Dash, "A Two-Step Deep Convolution Neural Network for Road Extraction from Aerial Images", 2019 6th International Conference On Signal Processing And Integrated Networks (SPIN), pp. 1-5

[4] Hussam Qassim, Abhishek Verma, David Feinzimer, "Compressed Residual-VGG16 CNN Model for Big Data Places Image Recognition", Image Recognition, 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1-7.

[5] Shikai Sun, Aoi Xia, Bingqi Zhang," Road Centerlines Extraction From High Resolution Remote Sensing Image, Igarss 2019 - 2019 Ieee International Geoscience And Remote Sensing Symposium, pp. 1-4.

[6] Zhong Qu 1,2, Jing Mei1, Ling Liu3, " Crack Detection Of Concrete Pavement With Cross-Entropy Loss Function And Improved VGG16 Network Model", IEEE Access (Volume: 8), Pp. 54564-54573.

[7] Anugrah Bintang Perdana, Adhi Prahara, "Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model", 2019 International Conference Of Computer Science And Information Technology (Icosnikom), pp. 1-4.

[8] Anil P.N., Dr. S. Natarajan, "A Novel Approach Using Active Contour Model for Semi-Automatic Road Extraction from High Resolution Satellite Imagery", 2010 Second International Conference on Machine Learning and Computing, pp. 1-4.

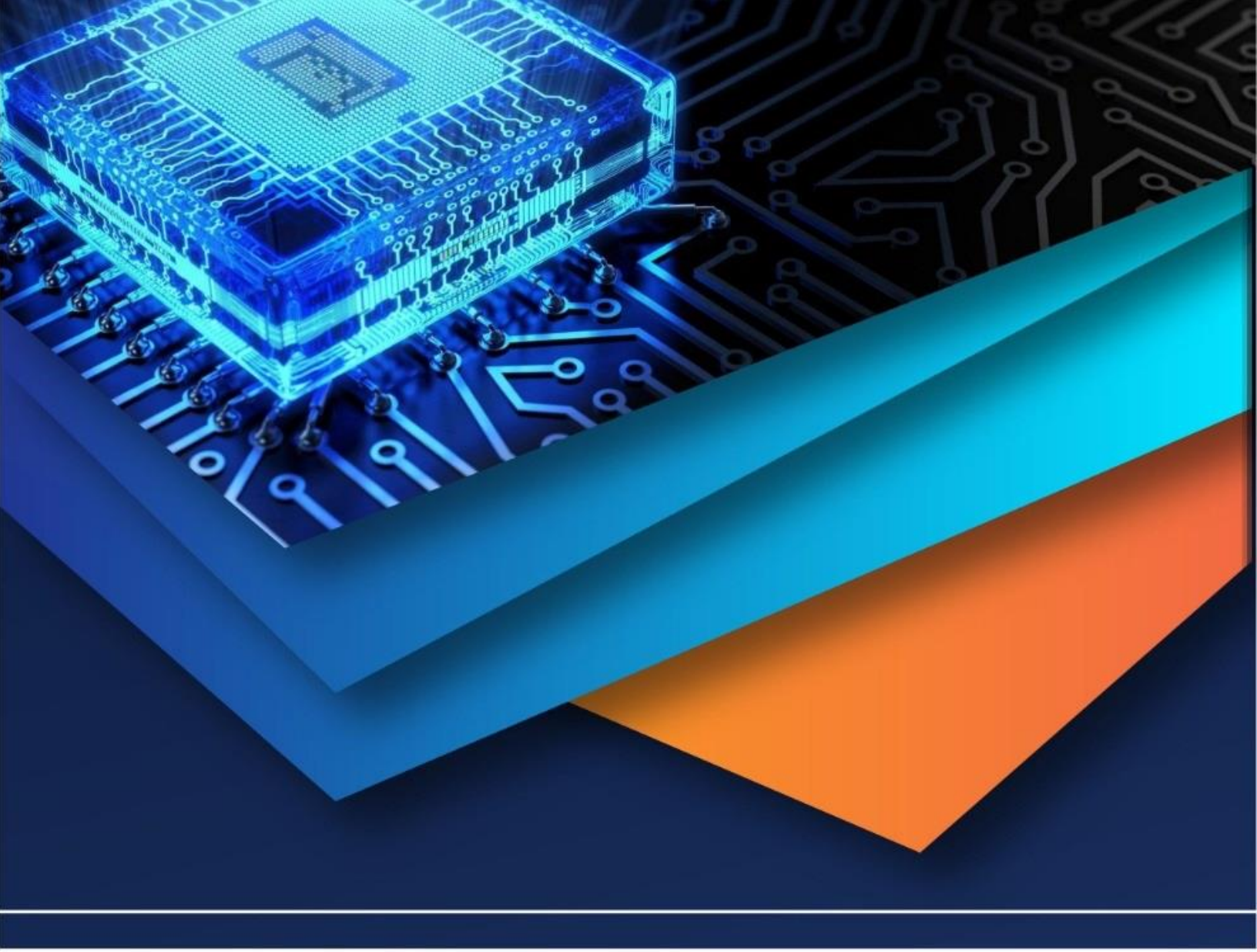
[9] Nontawat Pattanajak and Hossein Malekmohamadi, "Improving a 3-D Convolutional Neural Network Model Reinvented from VGG16 with Batch Normalization", 2019 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet Of People And Smart City Innovation (Smartworld/SCALCOM/UIC/ATC/Cbdcom/IOP/SCI), pp. 1-6.

[10] Xianxu Li, Qing Chang, And Yong Xu, "Queueing Characteristics Of The Best Effort Network Coding Strategy", 2019 22nd International Conference On Computer And Information Technology (ICCIT), pp. 1-8.

[11] Hashir Tanveer, Timo Balz, Bahaa Mohambdi, "Using Convolutional Neural network (CNN) Approach for Ship Detection in Sentunel-1 SAR Imagery", 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), pp. 1-5.

[12] Baolin Yang, Shixin Wang, Yi Zhou, "Extraction of road blockage information for the Jiuzhaigou earthquake based on a convolution neural network and very-high-resolution satellite images", Earth Science Informatics (2020), pp.1-13.

[13] A. Mohammadzadeh . M.J. Valadan Zoej . A. Tavakoli, "Automatic Main Road Extraction from High Resolution Satellite Imageries by Means of Particle Swarm Optimization Applied to a Fuzzy-based Mean Calculation Approach", Journal of the Indian Society of Remote Sensing, pp. 1-12.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY